

Slide 1 – Intro

- Today we'll be discussing our readings, discussing Hans Rosling & Gapminder, covering some basic data protocols and creating a web scraper
- I'll also be introducing the first assignment.

Slide 2 – Scrolling Forward

- Conversation starter this week is David Levy's "Scrolling Forward: Making Sense of Documents in the Digital Age"
- Levy is a fascinating character in 1981 he followed up his Ph.D. in computer science by studying bookbinding and calligraphy – text is an inciteful love letter to the printer word
- What were your thoughts on the selected reading? How does (might) it relate to data visualization?

Slide 3 – Receipt!

"The receipt is meant to function as 'proof of purchase,' as evidence that an exchange of money for goods actually took place. Coming into being at the very time and place the food was prepared and the goods were delivered, the receipt serves as witness to these facts. Its job is to tell its story in future situations, at other times and places — to play a role in other activities. It may be used for example, to return or exchange the items purchased, when requesting reimbursement of the cost of the purchase (when submitting a travel expense report, say) or as a way of documenting expenses for tax purposes." – Scrolling Forward, pg 14.

- Earlier in chapter there is mention of history of first accounting documents (Mesopotamian cuneiform tablets and Iranian records) are first instances of written words. Mention of scholar Denise Schmandt-Besserat who theorizes "writing was direct growth of counting and accounting practices conducted within increasingly complex, urban societies" (pg. 13)
- Receipt is repeatable, fixed, "ensures the story conveyed by the data it contains will be repeatedly available"
- Stable artifacts – vehicles of meaning

Slide 4 – Transcendental Data

- Alan Liu – a humanities scholar (with particular interest in Woodsworth) based at UCSB. Liu is a really provocative figure, he writes about data, culture and documents in a way that few others do. We'll return to him later in the semester with an excerpt from his text "The Laws of Cool"
- "Whether one writes fiction or business reports, prepares lectures or sales presentations, publishes works stored in a library or a commercial database..." an individual is producing a contemporary discourse – "encoding and structured markup" pg. 209

- Discussion about XML as a ubiquitous publishing protocol, but a metaphor that can be broadly applied
- Liu has a few questions: “How is an author now a postindustrial producer?” – which we aren’t so interested in. Also: “What are the aesthetics of encoded or structured discourse?”
- We’re not too interested in his discussion about how we might feed poetry to a database. Although it is interesting... using a relational database to sort by title, author, publication_date, line number and run queries etc...
- However, his breaking down poems into components is compelling, this idea we that could isolate elements like title, stanza or line, and wrap them within descriptive brackets
- Key points: “separation of content from material instantiation or formal presentation” pg. 216
- Vocabulary: “data-pours” live feeds and streams
- Let’s focus on pgs. 222-228

Questions:

Pg. 222 – What connection does Liu draw between Rifle manufacturing & document specifications? (refer to Liu’s poetry markup if necessary)

Pg. 221 – What is significance of terminology metastandardization and metamanagement?

Slide 6 – Rifle

- John Hall – forefather of industrialization. Preceded Frederick Winslow Taylor. (Who is Taylor/What is Taylorization?)
- Quoting Wendell Pieze (XML Developer/theorist) “developed a system by which guns could be made without the hand-crafting traditionally required of them... instead, the parts were all made to more-than-humanly possible close tolerances by machine, then assembled not by piece, but by type. That is any barrel could fit on any stock, with any receiver, any lock. Etc.” Pg. 223
- This required standards, jigs and gauges.
- Hop forward and Liu is talking about forms “A form, properly designed, enabled the office manager of that day to get the desired information recorded exactly in the shape he wanted it and without the use of personal instructions. Such a form might be regarded as analogous to the the ‘jig’ used in factories, which enabled a ‘machine hand’ to perform work which otherwise could have been done by a skilled mechanic” pg. 226
- Break between industrialism and postindustrialism is too shallow.

Slide 7 – IDH Ch. 5

- In terms of general methodology, many of you will want to get your hands on the IDH text as chapter 5 outlines some strategies for approaching information design.
- These include AIDA (Attention, Interest, Desire, Action) – a persuasive sales model, The Inverted Pyramid (Lead with questions, provide supporting information, provide details) and the Principle of Least Effort (relating to search, from information/library studies).

Slide 8 – LATCH

- Saul Wurman's LATCH
- Location – physical geography (i.e. medical texts based off physical anatomy)
- Alphabet – Letter sequence (i.e. dictionaries)
- Time – for understanding sequence/order
- Category – Taxonomy!
- Hierarchy – measure (food chains, animal class, size, etc.)

Pool group for multiple examples of each. Discuss previous precedents including Minard Map & The Evolution of the Web.

Slide 9 – Box score

Did/does anybody collect baseball cards/any current or former baseball fans?

Discussion of Baseball & stats

- Boxscores
- ERA
- Batting Average
- H, 2B, 3B, HR
- BB
- SO
- RBI
- SLG percentage $TB/AB [1b + (2 \times 2b) + (3 \times 3b) + (4 \times HR)/AB]$
- What else?

Slide 10 – Harold Baines

- **Introduce '21st century baseball card assignment'**

Slide 11 – Hans Rosling

- **How many of you know Hans Rosling?**
- Doctor, academic, statistician and advocate.

- Explored a range of fields including economic development, agriculture, poverty and health in Africa, Asia and Latin America and served as an advisor to the WHO and Unicef. One of his main talking points is that aid can't simply be distributed regionally as there is tremendous variety within regions, aid needs to be hyper-targeted.
- This research led him to start a non-profit called Gapminder Foundation – dedicated to visualizing global inequality. The organization developed several custom flash applications, at least one of which might be recognizable to you.
- We're going to watch a TED talk by Rosling now, where he demonstrates not only his software, but his talents as using data to tell stories:

Watch: <http://www.youtube.com/watch?v=hVimVzgtD6w>

- **Thoughts on this work?**
- Interestingly enough, Rosling's Gapminder software was purchased by Google in 2007 (shortly after he gave this talk) and is now available straight out of Google Spreadsheets (it is labelled as a 'motion chart') and it is accessible via the Google Visualization API as well.

Slide 12 – CSV

- Navigate to: <http://toronto.ca/open>
- <http://bit.ly/nrlr4Y> – download and open in text editor & spreadsheet application

Slide 13 – XML

- Navigate to <http://feeds.boingboing.net/boingboing/iBag> – show feedburner XML
- View source for closer look
- True example: http://www.w3schools.com/xml/cd_catalog.xml

Slide 14 – JSON

- Navigate to: <http://toronto.ca/open>
- <http://bit.ly/pJiQZx> – open in text editor